

AD-A186 335

PARAMETER ESTIMATION FOR THE DIRICHLET-MULTINOMIAL
DISTRIBUTION USING SUP (U) FLORIDA STATE UNIV
TALLAHASSEE DEPT OF STATISTICS P J DANAHER JUL 87
FSU-STATISTICS-M761 AFOSR-TR-87-1084 F/G 12/3

1/1

UNCLASSIFIED

NL

AD-A186 335

REPORT DOCUMENTATION PAGE

UNCLASSIFIED		1b. RESTRICTIVE MARKINGS		FILE COPY													
2a. SECURITY CLASSIFICATION AUTHORITY N/A		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited															
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A																	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Technical Report No. M-761		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-82-1084															
6a. NAME OF PERFORMING ORGANIZATION Florida State University	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM															
6c. ADDRESS (City, State and ZIP Code) Department of Statistics Tallahassee, FL 32306-5055		7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332-6448															
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0007 AFOSR No. 85-C-007															
6c. ADDRESS (City, State and ZIP Code) Building 410 Bolling, AFB, DC 20332-6448		10. SOURCE OF FUNDING NOS. <table border="1"><tr><td>PROGRAM ELEMENT NO. 6.1102F</td><td>PROJECT NO. 2304</td><td>TASK NO. AS</td><td>WORK UNIT NO.</td></tr></table>				PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304	TASK NO. AS	WORK UNIT NO.								
PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304	TASK NO. AS	WORK UNIT NO.														
11. TITLE (Include Security Classification) Parameter Estimation for the Dirichlet-Multinomial Distribution using Supplimentary Beta-Binomial Data																	
12. PERSONAL AUTHOR(S) Peter Panaher																	
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Yr., Mo., Day) July 1987		15. PAGE COUNT 12													
16. SUPPLEMENTARY NOTATION																	
17. COSATI CODES <table border="1"><tr><td>FIELD</td><td>GROUP</td><td>SUB GR</td></tr><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr></table>		FIELD	GROUP	SUB GR										18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) beta-binomial; Dirichlet-multinomial; magazine exposure; maximum likelihood estimation; method of moments.			
FIELD	GROUP	SUB GR															
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>We develop estimates for the parameters of the Dirichlet-multinomial distribution (DMD) when there is insufficient data to obtain maximum likelihood or method of moment estimates known in the literature. We do, however, have supplementary beta-binomial data pertaining to the marginals of the DMD, and use these data when estimating the DMD parameters. A real situation and data set are given where our estimates are applicable.</p>																	
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED															
22a. NAME OF RESPONSIBLE INDIVIDUAL F1 McJ Crawley M1		22b. TELEPHONE NUMBER (Include Area Code) 202-767-5025		22c. OFFICE SYMBOL AFOSR/NM													

AFOSR-TR- 87 - 1084

**Parameter Estimation for the Dirichlet-
Multinomial Distribution using
Supplementary Beta-Binomial Data**

by

Peter J. Danaher

Department of Statistics
Florida State University
Tallahassee, Florida 32306

FSU Technical Report M761
AFOSR Technical Report 87-209

July 1987

Research supported by the Air Force Office of Scientific
Research Grant AFOSR 85-C-0007.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



**DTIC
ELECTE
SEP 29 1987**
S D E

PARAMETER ESTIMATION FOR THE DIRICHLET-MULTINOMIAL DISTRIBUTION USING SUPPLEMENTARY BETA-BINOMIAL DATA

Peter J. Danaher
Department of Statistics
Florida State University
Tallahassee, Florida 32306

Key Words and Phrases: beta-binomial; Dirichlet-multinomial; magazine exposure; maximum likelihood estimation; method of moments.

ABSTRACT

We develop estimates for the parameters of the Dirichlet-multinomial distribution (DMD) when there is insufficient data to obtain maximum likelihood or method of moment estimates known in the literature. We do, however, have supplementary beta-binomial data pertaining to the marginals of the DMD, and use these data when estimating the DMD parameters. A real situation and data set are given where our estimates are applicable.

1. INTRODUCTION

Suppose we have $t + 1$ mutually exclusive events and Y_j is the number of times that event j occurs out of k independent trials, $j = 0, 1, \dots, t$. Let \vec{Y} , conditional on the vector of probabilities $\vec{\Pi} = \vec{\pi}$, have a multinomial distribution, i.e., $\vec{Y}|\vec{\Pi} = (Y_0, \dots, Y_t)|\vec{\Pi} = \vec{\pi} \sim \text{multinomial}(k, \pi_0, \dots, \pi_t)$. Let $\vec{\Pi}$ have a Dirichlet distribution; then compounding the multinomial distribution with the Dirichlet gives the so-called Dirichlet (or β -) compound multinomial distribution (Johnson and Kotz 1969), also known as the compound multinomial distribution (Mosimann 1962). It is commonly known as the Dirichlet-multinomial distribution, denoted by $\text{DMD}(k, \tau, \vec{\lambda})$, $\tau > 0$, $\lambda_j > 0$, $j = 0, 1, \dots, t$, $\sum_{j=0}^t \lambda_j = 1$.

An excellent literature review of the parameter estimation and applications of the DMD was given by Chuang and Cox (1985), although they did not mention the application of the DMD to magazine and TV exposure data (Chandon 1976; Leckenby and Kishi 1984; Rust and Leone 1984), an application we will give in Section 4.

The DMD mass function is

$$f^{DM}(\vec{Y} = \vec{y}) = \frac{k!}{(k - \sum_{j=1}^t y_j)!} \frac{\Gamma(\tau)}{\Gamma(\tau + k)} \frac{\Gamma(k - \sum_{j=1}^t y_j + \tau \lambda_0)}{\Gamma(\tau \lambda_0)} \times \prod_{j=1}^t \frac{\Gamma(y_j + \tau \lambda_j)}{\Gamma(\tau \lambda_j) y_j!}, \quad 0 \leq y_j \leq k, \quad j = 1, \dots, t, \quad \sum_{j=1}^t y_j \leq k, \quad (1.1)$$

where $\Gamma(l) = (l-1)\Gamma(l-1)$, the usual gamma function.

To fix ideas we will set $t = 3$. The data needed to estimate λ_j , $j = 0, 1, 2, 3$ is $(n_{i0}, n_{i1}, n_{i2}, n_{i3})$, $i = 1, \dots, n$, where n_{ij} is the number of occurrences of event j for the i^{th} person and n is the sample size ($\sum_{j=0}^3 n_{ij} = k$, $\forall i$). Denote the total number of people in the sample who fall into category j as $n_{.j} = \sum_{i=1}^n n_{ij}$.

Chuang and Cox (1985) estimated λ_j with

$$\hat{\lambda}_j = \frac{\bar{n}_{.j}}{k}, \quad (1.2)$$

where $\bar{n}_{.j} = n_{.j}/n$. We still need to find an estimate of τ , for which we now give four different estimates which have appeared in the literature.

Mosimann (1962) showed that the covariance matrix of $\vec{\Pi}$ (denoted $\Sigma_{\vec{\Pi}}$) and the covariance matrix of \vec{Y} (denoted $\Sigma_{\vec{Y}}$) are related thus,

$$\Sigma_{\vec{Y}} = \frac{k + \tau}{1 + \tau} \Sigma_{\vec{\Pi}}. \quad (1.3)$$

He suggested estimating $\Sigma_{\vec{\Pi}}$ with $\bar{n}_{.j}(k - \bar{n}_{.j})/k$ on the diagonal and $-\bar{n}_{.j}\bar{n}_{.j'}/k$, $j \neq j'$, on the off-diagonal and estimating $\Sigma_{\vec{Y}}$ with $\sum_{i=1}^n (n_{ij} - \bar{n}_{.j})^2/(n-1)$ on the diagonal and $\sum_{i=1}^n (n_{ij} - \bar{n}_{.j})(n_{ij'} - \bar{n}_{.j'})/(n-1)$ on the off-diagonal. Notice that $\hat{\Sigma}_{\vec{Y}}$ and $\hat{\Sigma}_{\vec{\Pi}}$ are nonsingular 3×3 matrices (Mosimann 1962). Then, using (1.3),

$$\frac{k + \hat{\tau}}{1 + \hat{\tau}} = \left(\frac{\det(\hat{\Sigma}_{\vec{Y}})}{\det(\hat{\Sigma}_{\vec{\Pi}})} \right)^{\frac{1}{3}},$$

from which $\hat{\tau}$ can be obtained.

Brier's (1980) estimate of τ similarly comes from solving for $\hat{\tau}$ in the following equation

$$\frac{k + \hat{\tau}}{1 + \hat{\tau}} = \frac{1}{3(n-1)} \sum_{i=1}^n \sum_{j=0}^3 \frac{(n_{ij} - \bar{n}_{.j})^2}{\bar{n}_{.j}}.$$

Both Mosimann's and Brier's estimates are based on the method of moments estimation technique. Owing to the form of Brier's estimate Chuang and Cox (1985) called it a chi-square moment estimate.

The likelihood equations used to find $\hat{\lambda}_j$ and $\hat{\tau}$ are

$$\begin{aligned} \sum_{i=1}^n \sum_{l=0}^{n_{ij}-1} \frac{1}{l + \lambda_j} &= \sum_{i=1}^n \sum_{l=0}^{n_{ij}-1} \frac{1}{l + \lambda_0}, \quad j = 1, 2, 3, \\ \lambda_0 &= 1 - \sum_{j=1}^3 \lambda_j, \\ \sum_{i=1}^n \sum_{j=0}^3 \sum_{l=0}^{n_{ij}-1} \frac{\lambda_j}{l + \tau \lambda_j} &= n \sum_{l=0}^{k-1} \frac{1}{l + \tau}. \end{aligned} \quad (1.4)$$

Due to the numerical difficulties of obtaining a solution to the likelihood equations of (1.4) Chuang and Cox (1985) estimated τ using the pseudo maximum likelihood method of Gong and Samenigo (1981). Chuang and Cox's method is to substitute the $\hat{\lambda}_j$ of (1.2) into (1.1) then obtain the likelihood equation which involves just the parameter τ . Their likelihood equation is

$$\sum_{i=1}^n \sum_{j=0}^3 \sum_{l=0}^{n_{ij}-1} \frac{\bar{n}_{.j}}{l + \bar{n}_{.j}\tau} = n \sum_{l=0}^{k-1} \frac{1}{l + \tau}.$$

However, if $k = 1$, we can use neither Mosimann's nor Brier's estimate of τ since $(k + \tau)/(1 + \tau)$ is 1 when $k = 1$. In addition, the maximum likelihood and pseudo maximum likelihood methods do not give unique solutions when $k = 1$ as there are only three linearly independent data and four parameters to estimate. It is precisely when $k = 1$ that we desire to estimate τ . A reason for this will be apparent in Section 4.

2. ESTIMATING τ WHEN $k = 1$

Let $X_1 = Y_1 + Y_3$ and $X_2 = Y_2 + Y_3$; then (X_1, X_2) is the bivariate distribution of the total number of occurrences of events 1 and 3 and events 2

and 3, respectively. The marginal distribution of each of the Y_j is the beta-binomial distribution (BBD) denoted by $BBD(k, \tau\lambda_j, \tau(1-\lambda_j))$, whose mass function is obtained by letting $t = 1$ in (1.1).

An application of some general DMD theorems in Basu and de B. Pereira (1982) shows that

$$\begin{aligned} X_1 &\sim BBD(k, \tau(\lambda_1 + \lambda_3), \tau(\lambda_0 + \lambda_2)), \\ X_2 &\sim BBD(k, \tau(\lambda_2 + \lambda_3), \tau(\lambda_0 + \lambda_1)). \end{aligned} \quad (2.1)$$

The joint mass function of X_1 and X_2 is

$$\begin{aligned} g(X_1 = x_1, X_2 = x_2) &= \frac{k!\Gamma(\tau)}{\Gamma(\tau + k)} \sum_{x_3 = \max\{0, x_1 + x_2 - k\}}^{\min\{x_1, x_2\}} \\ &\frac{\Gamma(x_1 - x_3 + \tau\lambda_1)\Gamma(x_2 - x_3 + \tau\lambda_2)\Gamma(x_3 + \tau\lambda_3)\Gamma(k + x_3 - x_1 - x_2 + \tau\lambda_0)}{(x_1 - x_3)!(x_2 - x_3)!x_3!(k + x_3 - x_1 - x_2)! \prod_{i=0}^3 \Gamma(\tau\lambda_i)}, \end{aligned} \quad (2.2)$$

$$0 \leq x_i \leq k, \quad i = 1, 2.$$

We saw in Section 1 that to estimate τ when $k = 1$ we need some extra data. From (2.1), $X_1 \sim BBD(k, \tau(\lambda_1 + \lambda_3), \tau(\lambda_0 + \lambda_2))$ so we can estimate $\tau(\lambda_1 + \lambda_3)$ and $\tau(\lambda_0 + \lambda_2)$ using supplementary data pertaining to X_1 , if such data is available; similarly for X_2 . Define α_i and β_i , $i = 1, 2$, as follows;

$$\begin{aligned} \tau(\lambda_1 + \lambda_3) &= \alpha_1, \quad \tau(\lambda_0 + \lambda_2) = \beta_1, \\ \tau(\lambda_2 + \lambda_3) &= \alpha_2, \quad \tau(\lambda_0 + \lambda_1) = \beta_2. \end{aligned} \quad (2.3)$$

From (2.3), $\alpha_i + \beta_i = \tau(\lambda_0 + \lambda_1 + \lambda_2 + \lambda_3) = \tau$, $i = 1, 2$. This means that when α_i and β_i are estimated using supplementary BBD data the estimates should be constrained so that

$$\alpha_1 + \beta_1 = \alpha_2 + \beta_2 = \tau. \quad (2.4)$$

The problem with trying to use constraint (2.4) is that τ is unknown. Chandon (1976) could not, so did not, apply constraint (2.4) when estimating α_i and β_i . As a result, $\hat{\alpha}_1 + \hat{\beta}_1 \neq \hat{\alpha}_2 + \hat{\beta}_2$ where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are MLEs or method of moment estimates obtained by using supplementary BBD data for X_i , $i = 1, 2$. Knowing this, he took a weighted average of $\hat{\alpha}_1 + \hat{\beta}_1$ and $\hat{\alpha}_2 + \hat{\beta}_2$ to estimate τ with

$$\hat{\tau}_c = \sum_{i=1}^2 (\hat{\alpha}_i + \hat{\beta}_i) \frac{w_i}{w_1 + w_2}, \quad \text{where } w_i = \frac{\hat{\alpha}_i}{\hat{\alpha}_i + \hat{\beta}_i}, \quad i = 1, 2. \quad (2.5)$$

We found this unappealing since this estimator of τ is rather *ad hoc*. He could equally well have chosen the arithmetic, geometric, or harmonic mean of $(\hat{\alpha}_i + \hat{\beta}_i)$ $i = 1, 2$.

Our procedure is as follows. Denote the correlation between X_1 and X_2 as ρ_{X_1, X_2} . Then (2.3) substituted into ρ_{X_1, X_2} gives

$$\begin{aligned}\rho_{X_1, X_2} &= \frac{\lambda_0 \lambda_3 - \lambda_1 \lambda_2}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_0 + \lambda_2)(\lambda_2 + \lambda_3)(\lambda_0 + \lambda_1)}} \\ &= \tau^2 \frac{(\lambda_0 \lambda_3 - \lambda_1 \lambda_2)}{\sqrt{\alpha_1 \beta_1 \alpha_2 \beta_2}}.\end{aligned}\quad (2.6)$$

Solving for τ in (2.6) gives

$$\tau = \left(\frac{\alpha_1 \beta_1 \alpha_2 \beta_2}{(\lambda_1 + \lambda_3)(\lambda_0 + \lambda_2)(\lambda_2 + \lambda_3)(\lambda_0 + \lambda_1)} \right)^{\frac{1}{2}}. \quad (2.7)$$

From (2.3) it follows that

$$\begin{aligned}\lambda_1 + \lambda_3 &= \frac{\alpha_1}{\alpha_1 + \beta_1}, & \lambda_0 + \lambda_2 &= \frac{\beta_1}{\alpha_1 + \beta_1}, \\ \lambda_2 + \lambda_3 &= \frac{\alpha_2}{\alpha_2 + \beta_2}, & \lambda_0 + \lambda_1 &= \frac{\beta_2}{\alpha_2 + \beta_2}.\end{aligned}\quad (2.8)$$

Substituting the four equations of (2.8) into (2.7) gives

$$\tau = \sqrt{(\alpha_1 + \beta_1)(\alpha_2 + \beta_2)}. \quad (2.9)$$

The above construction shows that it is more reasonable to estimate τ with the geometric mean of $\hat{\alpha}_i + \hat{\beta}_i$, i.e., $\hat{\tau}_{gm} = \sqrt{(\hat{\alpha}_1 + \hat{\beta}_1)(\hat{\alpha}_2 + \hat{\beta}_2)}$, rather than the weighted average estimate, $\hat{\tau}_c$.

3. ASYMPTOTIC PROPERTIES OF $\hat{\tau}$

If α_i and β_i are estimated with consistent estimates then, as $n \rightarrow \infty$,

$$\hat{\tau}_c \rightarrow \frac{\alpha_1 + \alpha_2}{\frac{\alpha_1}{\alpha_1 + \beta_1} + \frac{\alpha_2}{\alpha_2 + \beta_2}},$$

which equals τ iff (2.4) holds. On the other hand, $\hat{\tau}_{gm} \rightarrow \tau$ iff (2.9) holds. Since (2.4) \Rightarrow (2.9), but the converse is not true, $\hat{\tau}_{gm}$ is consistent under a weaker assumption than that required to make $\hat{\tau}_c$ consistent.

We can compare the asymptotic relative efficiency (*ARE*) of $\hat{\tau}_c$ and $\hat{\tau}_{gm}$ by examining the ratio of their asymptotic variances. Define $ARE = AV(\hat{\tau}_c)/AV(\hat{\tau}_{gm})$ where $AV(\cdot)$ denotes asymptotic variance. The *AV* of the two competing estimates can be obtained from knowledge of the asymptotic joint distribution of $(\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)$ and use of the delta method.

If the $\hat{\alpha}_i$'s and $\hat{\beta}_i$'s are MLEs then the asymptotic joint distribution can be obtained from the from general MLE theory (Lehmann 1983), i.e.,

$$\sqrt{n}(\tilde{\theta} - \bar{\theta}) \rightarrow MVN(0, I^{-1}(\bar{\theta})) \quad , \text{ as } n \rightarrow \infty \quad , \quad (3.1)$$

where $\bar{\theta}' = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ and $I(\bar{\theta})$ is the information matrix. Danaher (1987) proved that the regularity conditions for (3.1) to be true are satisfied for the MLEs of the BBD parameters.

Denote the information matrix of (α_i, β_i) as $I(\alpha_i, \beta_i)$, and the mass function of $X_i \sim BBD(k, \alpha_i, \beta_i)$ as $f_{x_i}^{BBD}$, $i = 1, 2$. Then

$$I(\alpha_i, \beta_i) = \begin{bmatrix} \sum_{x_i=1}^k \Delta(\alpha_i, x_i) f_{x_i}^{BBD} - \Delta(\alpha_i + \beta_i, k) & , -\Delta(\alpha_i + \beta_i, k) \\ -\Delta(\alpha_i + \beta_i, k) & , \sum_{x_i=0}^{k-1} \Delta(\beta_i, k - x_i) f_{x_i}^{BBD} - \Delta(\alpha_i + \beta_i, k) \end{bmatrix} \quad ,$$

where $\Delta(\gamma, l) = \sum_{j=0}^{l-1} 1/(\gamma + j)^2$.

If it is assumed that the bivariate distribution $(\hat{\alpha}_1, \hat{\beta}_1)$ is independent of $(\hat{\alpha}_2, \hat{\beta}_2)$ then

$$I(\bar{\theta}) = \begin{bmatrix} I(\alpha_1, \beta_1) & 0 \\ 0 & I(\alpha_2, \beta_2) \end{bmatrix} \quad .$$

To ensure that Lehmann's (1983, p345) definition of asymptotic relative efficiency is well defined, we must assume that (2.4) is true when comparing the asymptotic variances of $\hat{\tau}_c$ and $\hat{\tau}_{gm}$. Use of the delta method and (2.4) gives

$$ARE = \bar{x}_c' I^{-1}(\bar{\theta}) \bar{x}_c / (\bar{x}_{gm}' I^{-1}(\bar{\theta}) \bar{x}_{gm}) \quad ,$$

where $\bar{x}_c' = \frac{1}{\alpha_1 + \alpha_2} (\alpha_1, \alpha_1, \alpha_2, \alpha_2)$ and $\bar{x}_{gm}' = \frac{1}{2} (1, 1, 1, 1)$.

Clearly, when $\alpha_1 = \alpha_2$ the $ARE = 1$. Some ARE s for selected α_i 's and β_i 's are given in Table I. The table shows that the ARE is greater than one for three of the four cases considered. An interesting observation is that for given α_i and β_i the ARE does not vary much with k . Due to the complexity

Table I: *ARE* comparison of $\hat{\tau}_c$ and $\hat{\tau}_{gm}$ for some α_i 's and β_i 's.

k	α_i and β_i			
	$\alpha_1 = 1 \beta_1 = 2$	$\alpha_1 = 0.5 \beta_1 = 1$	$\alpha_1 = 0.1 \beta_1 = 0.4$	$\alpha_1 = 10 \beta_1 = 5$
	$\alpha_2 = 2 \beta_2 = 1$	$\alpha_2 = 0.2 \beta_2 = 1.3$	$\alpha_2 = 0.3 \beta_2 = 0.3$	$\alpha_2 = 3 \beta_2 = 12$
2	1.11	0.93	1.02	1.24
4	1.11	0.92	1.01	1.24
8	1.11	0.91	1.01	1.24

of the information matrix the author was unable to find conditions on α_i and β_i under which $ARE > 1$. Hence, to check the conjecture that $ARE > 1$ most of the time, two hundred randomly chosen α_i 's and β_i 's were selected to conform to (2.4) and the *ARE* calculated. Some $144/200 = 72\%$ of the cases had $ARE > 1$. Hence $\hat{\tau}_{gm}$ is asymptotically more efficient than $\hat{\tau}_c$ for approximately three-quarters of the possible α_i and β_i which satisfy (2.4).

4. APPLICATIONS

Suppose an advertiser is about to launch an advertising campaign by placing k ads in each of two different magazines. To evaluate the effectiveness of the campaign the advertiser would like to estimate the proportion of the population which sees at least one of the ads (known as the *reach*). We say that a person is *exposed* to an ad when he or she sees the ad. Let Y_1 = the number of exposures *exclusive* to magazine 1, Y_2 = the number of exposures *exclusive* to magazine 2, and Y_3 = the number of exposures to *both* magazines 1 and 2, for a particular person with $0 \leq Y_i \leq k$, $i = 1, 2, 3$. Then X_i , as defined in Section 2, is the number of exposures a person has to magazine i , $i = 1, 2$. Chandon (1976) modelled \vec{Y} with the DMD and the exposure distribution for a single magazine (X_i here) was first modelled with the BBD by Metheringham (1964). The DMD has also been used to model combined TV and magazine exposure data (Rust and Leone 1984).

In the media survey we used for our data two questions were asked of the respondents (for weekly magazines);

Q1) "Have you personally read or looked into any issue of ... (magazine name) in the last seven days - it doesn't matter where?" (Has a Y/N answer).

Q2) "How many different issues of ...(magazine name), if any, do you personally read or look into in an average month - it doesn't matter where?" (Has answer 0,1,2,3,4 issues).

The wording of Q1 and Q2 are modified appropriately for two-weekly, monthly and two-monthly magazines. These questions were asked for forty different magazines.

An implicit assumption in the magazine advertising field is that a person who reads a magazine is exposed to all the advertisements in that magazine. This is unlikely to be true for people who meet the criterion of "read" in Q1 and Q2. However it is usually impractical to ask respondents which advertisements they have been exposed to so we cannot avoid making this assumption for the available data.

There are many media schedules an advertising agency can specify whose exposure distribution cannot be directly estimated from Q1 and Q2. For example, a schedule with 3 ads in each of two different magazines cannot be estimated using Q1 and Q2. We want, therefore, to construct a model which not only estimates observable exposure distributions accurately but can be used to estimate (or predict) exposure distributions outside the range of exposure distributions covered by Q1 and Q2.

The first step in using (1.1) to model the exposure distribution of \bar{Y} is to estimate the parameters of the DMD. When solving the likelihood equations we come up against a data problem. The response to Q2 does not tell us what a person's reading behavior was in a specified week, it only gives us the total number of issues read in the last four weeks. If a person's response to Q2 is "4" then clearly they saw each issue but if their response is a "2" we have no way of knowing which two issues were read. Hence Q2 cannot give us the data required to fit (1.1). We can, however, use Q1 to fit (1.1) because here a Yes/No response tells us precisely whether or not a person read the last issue of a magazine. The problem this time is that $k = 1$ when using Q1 to fit the DMD and we saw in Section 1 that the conventional methods for estimating τ do not work when $k = 1$. We can, therefore, use (2.9) to estimate τ in the following way.

The response to Q1 gives us data $(n_{i0}, n_{i1}, n_{i2}, n_{i3}), \sum_j n_{ij} = 1, n_{ij} \in \{0, 1\} \forall i$. These data can be used to estimate λ_j using (1.2). The response to Q2 gives us data $\{n_{x_i}\}, i = 1, 2$ where n_{x_i} is the number of people in the sample who have x_i exposures to magazine $i, 0 \leq x_i \leq 4, i = 1, 2$. Hence we

Table II: Observed Exposure Distributions and Parameter Estimates for the *New Zealand Listener (NZL)* and *Time Magazine*; $n = 5201$.

Exposures	Observed Univariate Exposure Distribution				
	0	1	2	3	4
<i>NZL</i>	2741	322	286	94	1758
<i>Time</i>	4373	301	186	54	287
Bivariate Data	$(n_{.0}, n_{.1}, n_{.2}, n_{.3}) = (2975, 1741, 189, 296)$ where $n_{.j} = \sum_i n_{ij}$				
Parameter Estimates	$\hat{\alpha}_1 = 0.0743$	$\hat{\beta}_1 = 0.1103$	$\hat{\tau}_c = 0.2473$		
	$\hat{\alpha}_2 = 0.0498$	$\hat{\beta}_2 = 0.4517$	$\hat{\tau}_{gm} = 0.3043$		

can use $\{n_{x_i}\}$ to get estimates of α_i and β_i in (2.3) by using the method of moments or maximum likelihood estimation for the BBD.

Once the parameters of (2.2) have been estimated we can estimate (or predict) the mass function of (X_1, X_2) for values of k other than 1 or 4, the values available from the data.

In Table II we give the observed univariate and bivariate exposure distributions for the *New Zealand Listener* and *Time Magazine* along with the MLEs for α_i and β_i . Since (2.4) does not hold (even approximately) for the parameter estimates we cannot use our derived form of the *ARE* to compare the asymptotic efficiency of $\hat{\tau}_c$ and $\hat{\tau}_{gm}$.

The exposure distribution of interest to advertisers is not the bivariate exposure distribution (X_1, X_2) , but rather $X_{tot} = X_1 + X_2$, i.e., the total number of exposures a person has to the ad campaign. Having estimated the parameters of (2.2) we obtain an estimate of the probability mass function of X_{tot} by a change of variables. If we denote the mass function of the exposure distribution as $f(X_{tot})$ then reach is $1 - f(X_{tot} = 0)$. The observed reach is 52.5% while the estimated reaches using $\hat{\tau}_{gm}$ and $\hat{\tau}_c$ are, respectively, 53.1% and 51.5%. Hence use of $\hat{\tau}_{gm}$ in (2.2) gives a closer estimate of reach than when $\hat{\tau}_c$ is used.

To further demonstrate the usefulness of the geometric mean estimate of τ we consider the data from Mosimann (1962). These data are frequencies of occurrence of different pollen grains made at $n = 73$ different core levels. Here the pollen counts totalled 100 at each core, i.e., $k = 100$ so there

Table III: Six Estimates of τ for Mosimann's (1962) pollen data.

Mosimann's	Brier's	pseudo-MLE	MLE	$\hat{\tau}_{gm}$	$\hat{\tau}_c$
81.92	73.21	62.97	60.19	57.76	54.23

is no need to use $\hat{\tau}_{gm}$ as the four methods outlined in Section 1 are all applicable, with maximum likelihood being the best (Chuang and Cox 1985). Nonetheless, we will estimate τ with $\hat{\tau}_{gm}$ to show that it competes admirably with the four estimates in Section 1.

Let Y_0 =pine, Y_1 =oak, Y_2 =alder and Y_3 =fir pollen counts (cf. Mosimann (1962) for details of these data).

Table III gives the estimates of τ using the four techniques in Section 1 as well as for $\hat{\tau}_{gm}$ and $\hat{\tau}_c$ when α_i and β_i are estimated by maximum likelihood. The estimate which is closest to the MLE is $\hat{\tau}_{gm}$, even closer to the MLE than the pseudo-MLE. Chuang and Cox (1985) point out that the pseudo-MLE is both easier to calculate and asymptotically comparable to the MLE. Their estimate does require some degree of programming, however, as do the MLEs of α_i and β_i used to calculate $\hat{\tau}_{gm}$. If $\hat{\alpha}_i$ and $\hat{\beta}_i$ for $\hat{\tau}_{gm}$ are estimated by the method of moments the computations required can easily be conducted on a calculator. Estimating α_i and β_i by the method of moments for Mosimann's data gives $\hat{\tau}_{gm} = 56.69$. This estimate is still quite close to the MLE estimate of τ in Table III and has the advantage of requiring no programming whatsoever.

ACKNOWLEDGEMENT

The author would like to thank Professor Duane Meeter for discussions which helped in the evolution of the estimator examined in this study. The author would also like to thank AGB:McNair Surveys New Zealand Ltd. for providing the magazine exposure data.

REFERENCES

- Basu, D. and de B. Pereira, C. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse. *J. of Statistical Planning and Inference*. 6, 345-362.
- Brier, S.S. (1980). Analysis of contingency tables under cluster sampling. *Biometrika*. 67, 591-596.
- Chandon, J-L. J. (1976). A comparative study of media exposure models. Ph.D. dissertation, Northwestern University.
- Danaher, P.J. (1987). Estimating multidimensional tables from survey data: predicting magazine audiences. Ph.D dissertation, Florida State University.
- Chuang, C. and Cox, C. (1985). Pseudo maximum likelihood estimation for the Dirichlet-multinomial distribution. *Commun. Statist. (Theor. Meth.)*. 14(10), 2293-2311.
- Gong, G. and Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* 9, 861-869.
- Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions*. John Wiley and Sons, New York.
- Leckenby, J.D. and Kishi, S. (1984). The Dirichlet-multinomial distribution as a magazine exposure model. *J. of Marketing Research*. 21, 100-106.
- Lehmann, E.L. (1983). *Theory of Point Estimation*. John Wiley and Sons, New York.
- Metherringham, R.A. (1964). Measuring the net cumulative coverage of a print campaign. *J. of Advertising Research* 4, 23-28.
- Mosimann, J.E. (1962). On the multinomial distribution, the multivariate β -distribution and correlations among proportions. *Biometrika*, 49, 65-82.
- Rust, R.T. and Leone, R.P. (1984). The mixed media Dirichlet-multinomial distribution: a model for evaluating television-magazine advertising schedules. *J. of Marketing Research*. 21, 89-99.

END

DATE

FILMED

JAN

1988